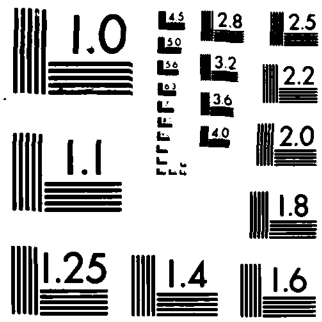
AD-A174 517 ON THE MEAN SQUARED ERROR OF NONPARAMETRIC QUANTILE ESTIMATORS UNDER RAND..(U) SOUTH CAROLINA UNIV COLUMBIA DEPT OF STATISTICS Y L LIO ET AL. SEP 86 TR-122 1/1

UNCLASSIFIED AFOSR-TR-86-2055 AFOSR-84-0156 F/G 12/1 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6

ON THE MEAN SQUARED ERROR OF NONPARAMETRIC QUANTILE ESTIMATORS UNDER RANDOM RIGHT-CENSORSHIP*

by

Y. L. Lio and W. J. Padgett

University of South Carolina
Statistics Technical Report No. 122
62G05-17

DEPARTMENT OF STATISTICS

The University of South Carolina
Columbia, South Carolina 29208

ON THE MEAN SQUARED ERROR OF NONPARAMETRIC QUANTILE ESTIMATORS UNDER RANDOM RIGHT-CENSORSHIP*

by

Y. L. Lio and W. J. Padgett

University of South Carolina
Statistics Technical Report No. 122
62G05-17

September, 1986

Department of Statistics
University of South Carolina
Columbia, SC 29208

*
Research supported by the U.S. Air Force Office of Scientific Research grant number AFOSR-84-0156 and U.S. Army Research office grant number MIPR ARO 139-85.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Stat. Tech. Rep. No. 122	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR. 86-2055
6a. NAME OF PERFORMING ORGANIZATION Department of Statistics 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code) University of South Carolina Columbia, SC 29208	7b. ADDRESS (City, State and ZIP Code) Directorate of Mathematics and Information Sciences Bolling AFB, DC 20332
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-84-0156
8c. ADDRESS (City, State and ZIP Code) Bolling AFB, DC 20332	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO.: 61102F; PROJECT NO.: 2304; TASK NO.: A5; WORK UNIT NO.:

11. TITLE (Include Security Classification)
On the Mean Squared Error of Nonparametric Quantile Estimators Under Random Right-Censorship

12. PERSONAL AUTHOR(S)
Y. L. Lio and W. J. Padgett

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM ____ TO ____	September, 1986	11

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Censored data; Kernel estimation; Product-limit quantile function; Mean-squared error; Large sample size

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

For randomly right-censored data, new asymptotic expressions for the mean squared errors of the product-limit quantile estimator and a kernel-type quantile estimator are presented in this paper. From these results a comparison of the two quantile estimators with respect to their mean squared errors is given.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian W. Woodruff	22b. TELEPHONE NUMBER (Include Area Code) (202) 767-5027 22c. OFFICE SYMBOL NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE. UNCLASSIFIED

ON THE MEAN SQUARED ERROR OF NONPARAMETRIC QUANTILE ESTIMATORS UNDER RANDOM RIGHT-CENSORSHIP

Y. L. Lio and W. J. Padgett

Department of Statistics
University of South Carolina
Columbia, South Carolina 29208

ABSTRACT

For randomly right-censored data, new asymptotic expressions for the mean squared errors of the product-limit quantile estimator and a kernel-type quantile estimator are presented in this paper. From these results a comparison of the two quantile estimators with respect to their mean squared errors is given.

1. INTRODUCTION

One characteristic of a probability distribution function that is of interest in many situations is the quantile function. The quantile function of a distribution function G is defined by $Q(p) = G^{-1}(p) = \inf\{x: G(x) \geq p\}$, $0 \leq p \leq 1$. For a random (uncensored) sample from G, the sample quantile function, $G_n^{-1}(p) = \inf\{x: G_n(x) \geq p\}$, has been used as a nonparametric estimator of $Q(p)$, where G_n denotes the sample distribution function. Many of the known results about $G_n^{-1}(p)$ have been presented by Csörgo (1983). In addition, Falk (1984, 1985) has studied the mean squared errors of the sample quantile and

kernel-type estimators and obtained asymptotic normality results for kernel estimators. Yang (1985) has obtained some convergence properties of kernel estimators of $Q(p)$ and has presented some simulation results comparing kernel-type estimators with other estimators.

For right-censored data, Sander (1975) discussed the estimation of $Q(p)$ by the quantile function of the product-limit estimator. She and Cheng (1984) derived asymptotic properties and Csörgő (1983) presented strong approximation results for that estimator.

For randomly right-censored data, Padgett (1986) discussed a smooth nonparametric estimator of the quantile function, defined by $Q_n(p) = h^{-1} \int_0^1 \hat{Q}_n(t)\, K((t-p)/h)\, dt$, where $\hat{Q}_n$ denotes the product-limit quantile function, K is a kernel function, and h is the bandwidth. This estimator, which had been mentioned briefly by Parzen (1979), was shown to be strongly consistent, and Q_n and an approximation, Q_n^*, were shown to be almost surely asymptotically equivalent. The asymptotic normality of Q_n and Q_n^* and some asymptotic mean equivalence and mean square convergence results were obtained by Lio, Padgett and Yu (1986) and Lio and Padgett (1986). Some simulation results in Padgett (1986) showed that, for exponential life and censoring distributions for fixed n and p, there were values of h for which the mean squared errors of $Q_n(p)$ were smaller than those of $\hat{Q}_n(p)$. More extensive simulations by Padgett and Thombs (1986) indicated the same results for several families of life distributions, kernel functions, and censoring distributions.

In this paper, new asymptotic expressions for the mean squared errors of $\hat{Q}_n(p)$ and $Q_n(p)$ are derived. The conditions on $Q_n(p)$ here are less restrictive than those required for the mean square convergence results of Lio and Padgett (1986). The expressions provide a comparison of the mean squared errors of these two estimators for small h and large n. In Section 2 some further notation and definitions are presented. The asymptotic expression for the mean squared error of the product-limit

quantile function is given in Section 3, and the result for the kernel estimator Q_n is derived in Section 4. It should be mentioned that the order statistic methods used by Falk (1984, 1985) to obtain an asymptotic expression for the mean squared error of the empirical quantile function cannot be used to study the mean squared error of the product-limit quantile function due to the unequal random jumps in the product-limit distribution function.

2. NOTATION AND PRELIMINARIES

Let $X_1^o, \ldots, X_n^o$ denote the true survival times of n items or individuals that are censored on the right by a sequence U_1, U_2, ..., U_n which in general may be either constants or random variables. It is assumed that the X_i^o's are nonnegative independent identically distributed random variables with common unknown distribution function F_o and unknown quantile function $Q^o = F_o^{-1}$. The observed right-censored data are denoted by the pairs (X_i, Δ_i), i=1,..., n where

$$X_i = \min\{X_i^o, U_i\}, \qquad \Delta_i = \begin{cases} 1 \text{ if } X_i^o \le U_i \\ \\ 0 \text{ if } X_i^o > U_i . \end{cases}$$

Let (Z_i, Λ_i), i=1,...,n, denote the ordered X_i's along with their corresponding Δ_i's. A popular estimator of the survival function $S_o = 1-F_o$ is the product-limit estimator of Kaplan and Meier (1958), shown to be "self-consistent" by Efron (1967) and defined by

$$\hat{P}_n(t) = \begin{cases} 1, & 0 \le t \le Z_1, \\ \prod_{i=1}^{k-1} \left(\frac{n-i}{n-i+1}\right)^{\Lambda_i}, & Z_{k-1} < t \le Z_k, \quad k=2,\ldots,n \\ 0, & t > Z_n . \end{cases}$$

Denote the product-limit estimator of $F_o(t)$ by $\hat{F}_n(t) = 1 - \hat{P}_n(t)$, and let s_j denote the jump of $\hat{P}_n$ at Z_j, that is,

$$s_j = \begin{cases} 1 - \hat{P}_n(Z_2), & j = 1 \\ \hat{P}_n(Z_j) - \hat{P}_n(Z_{j+1}), & j = 2,\ldots,n-1 \\ \hat{P}_n(Z_n), & j = n. \end{cases}$$

Note that $s_j=0$ if and only if $\Lambda_j = 0$, $j < n$, i.e. whenever Z_j is a censored observation. Also, denote $S_i \equiv \hat{F}_n(Z_{i+1}) = \sum_{j=1}^{i} s_j$, $i=1,\ldots,n$, with $S_0 \equiv 0$, $Z_0 \equiv 0$, and $Z_{n+1} \equiv Z_n + \varepsilon$, for some positive constant ε.

It is natural to estimate $Q^o(p)$ by the product-limit (PL) quantile function $\hat{Q}_n(p) = \inf\{t: \hat{F}_n(t) \geq p\}$. The kernel-type estimator $Q_n(p)$ studied by Padgett (1986) is written as

$$\begin{aligned} Q_n(p) &= h^{-1} \int_0^1 \hat{Q}_n(t)K((t-p)/h)dt \\ &= h^{-1} \sum_{i=1}^{n} Z_i \int_{S_{i-1}}^{S_i} K((t-p)/h)dt, \qquad (2.1) \end{aligned}$$

for a kernel function K and bandwidth h.

For the results here, the random right-censorship model will be assumed; that is, $U_1,\ldots,U_n$ constitute a random sample from a distribution H (usually unknown) and are independent of $X_1^o,\ldots, X_n^o$. The distribution function of each X_i, $i=1,\ldots,n$, is then $F = 1 - (1-F_o)(1-H)$.

For a distribution function G, let $T_G = \sup\{t: G(t)<1\}$.

3. MEAN SQUARED ERROR OF THE PL QUANTILE FUNCTION

In this section, an asymptotic expression for the mean squared error of the PL quantile function is derived. In the proof of this result, $K^*(t,s)$ denotes the generalized Kiefer process (cf. Csörgő, 1983, p. 118).

Theorem 3.1 Let p be such that $0 \leq p < \min\{1, T_{H(Q^o)}\}$. Suppose H is continuous and Q^o is twice differentiable in a neighborhood of p with bounded second derivative on a neighborhood of p. Then for large n, $E\{[\hat{Q}_n(p)-Q^o(p)]^2\}$ exists and

$$E\{[\hat{Q}_n(p)-Q^o(p)]^2\} = n^{-1} (Q^{o\prime}(p))^2 (1-p)^2 \int_0^p \frac{dx}{(1-x)^2(1-H(Q^o(x)))} + O(n^{-3/2}) + o(n^{-1}). \qquad (3.1)$$

<u>Proof:</u> Denoting the PL quantile function based on the uniform distribution on (0,1) by $U_n(p)$, we have $E\{[\hat{Q}_n(p)]^2\} = E\{[F_o^{-1}(U_n(p))]^2\}$. By Aly, Csörgo and Horváth (1985), $U_n(p) \le p^*$ a.s. if $p < p^* < \min\{1, T_{H(Q^o)}\}$ so that $F_o^{-1}(U_n(p)) \le F_o^{-1}(p^*)$ a.s. Hence, $E\{[\hat{Q}_n(p) - Q^o(p)]^2\} < \infty$.

Next, define the events $A_n = \{|U_n(p)-p|>\varepsilon\}$ for fixed $\varepsilon > 0$. Then

$$E\{[\hat{Q}_n(p) - Q^o(p)]^2\} = E\{[\hat{Q}_n(p)-Q^o(p)]^2 I_{A_n}\} + E\{[\hat{Q}_n(p)-Q^o(p)]^2 I_{A_n^c}\}, \qquad (3.2)$$

where I_A denotes the indicator random variable of the event A. By Földes and Rejtö (1981) and the symmetry property as in Sander (1975), $\varepsilon>0$ can be chosen so that $P[|U_n(p)-p|>\varepsilon] \le d_o \exp(-nd_1)$ for some positive constants d_o and d_1 where d_o does not depend on F_o and H. Then from (3.2)

$$E\{[\hat{Q}_n(p)-Q^o(p)]^2\} = O(\exp(-nc)) + E\{[\hat{Q}_n(p)-Q^o(p)]^2 I_{A_n^c}\} \qquad (3.3)$$

for some constant $c > 0$.

Now the second term on the right side of (3.3) is

$$E\{[F_o^{-1}(U_n(p))-F_o^{-1}(p)]^2 I_{A_n^c}\}$$
$$= E\{[Q^{o\prime}(p)\ (U_n(p)-p) + Q^{o\prime\prime}(p_1)\ (U_n(p)-p)^2/2]^2 I_{A_n^c}\} \qquad (3.4)$$

where p_1 belongs to a neighborhood of p. But (3.4) is equal to

$$E\{ Q^{o\prime}(p)(U_n(p)-p)^2\ I_{A_n^c}\} + O(E[|U_n(p)-p|^3])$$
$$= (Q^{o\prime}(p))^2\ E[(U_n(p)-p)^2] + O(n^{-3/2})$$

since

$$n^{3/2}|U_n(p)-p|^3 \le n^{3/2}\sup_{0\le p\le p^*}(|U_n(p)-p|^3)$$
$$\le n^{3/2}\sup_{0\le p\le p^{**}}(|\alpha_n(p)-p|^3),$$

the last term of which is uniformly integrable, where $p^* < p^{**} < \min\{1, T_{H(Q^o)}\}$ and α_n is the PL empirical quantile function based on the uniform distribution over (0,1).

So

$$E\{[\hat{Q}_n(p)-Q^o(p)]^2\} = (Q^{o\prime}(p))^2 n^{-1} E\{[n^{1/2}(U_n(p)-p)$$

$$-n^{-\frac12}K^*(p,n)]^2 + 2\,[n^{\frac12}(U_n(p)-p)$$
$$-n^{-\frac12}K^*(p,n)]n^{-\frac12}K^*(p,n)$$
$$+[n^{-\frac12}K^*(p,n)]^2\} = O(n^{-3/2})$$
$$= (Q^{o\prime}(p))^2 n^{-1} E\{[n^{\frac12}(U_n(p)-p)-n^{-\frac12}K^*(p,n)]^2\}$$
$$+(Q^{o\prime}(p))^2 n^{-1} E\{2[n(U_n(p)-p)-n^{-\frac12}K^*(p,n)]n^{-\frac12}K^*(p,n)\}$$
$$+(Q^{o\prime}(p))^2(1-p)^2 n^{-1}\int_0^p \frac{dx}{(1-x)^2(1-H(Q^o(x)))} + O(n^{-3/2}).$$

The result of the theorem follows from the facts that $E\{[n^{\frac12}(U_n(p)-p)-n^{-\frac12}K^*(p,n)]^2\} < \infty$ and $E\{[n^{\frac12}(U_n(p)-p)-n^{-\frac12}K^*(p,n)]n^{-\frac12}K^*(p,n)\} < \infty$, since $[n^{\frac12}(U_n(p)-p)-n^{-\frac12}K^*(p,n)]^r$ is uniformly integrable for $r \geq 1$ and $n^{\frac12}(U_n(p)-p)-n^{-\frac12}K^*(p,n)\to 0$ a.s. as $n\to\infty$.///

4. MEAN SQUARED ERROR OF THE KERNEL ESTIMATOR

The mean squared error of the kernel quantile estimator $Q_n(p)$ is considered in this section. Theorem 4.1 gives the asymptotic expression.

Theorem 4.1 Let p be such that $0 \leq p < \min\{1, T_{H(Q^o)}\}$. Suppose H is continuous, Q^o is twice differentiable in a neighborhood of p with bounded second derivative, and $Q^{o\prime}(p) > 0$. Assume that the kernel K has support $[-c,c]$ and $\int K(x)dx = 1$ and $\int x\,K(x)dx = 0$ for some $c > 0$. Then

$$E\{[Q_n(p)-Q^o(p)]^2\} = n^{-1}(Q^{o\prime}(p))^2(1-p)^2\int_0^p \frac{dx}{(1-x)^2[1-H(Q^o(x))]}$$
$$+\,2n^{-1}(1-p)^2(Q^{o\prime}(p))^2\int_{-c}^{c} K(t)[1-\tilde{K}(t)]\int_p^{p+ht}\frac{dx}{(1-x)^2[1-H(Q^o(x))]}\,dt$$
$$+O(n^{-3/2})+O(h^2)+O(h^2n^{-1})+o(hn^{-1})+o(n^{-1}),$$

where $\tilde{K}(t) = \int_{-c}^{t} K(x)dx$ for $-c \leq t \leq c$.

Proof: First, write

$$E\{\,[Q_n(p)-Q^o(p)]^2\} = E\{[\int_{-c}^{c}(\hat{Q}_n(p+hu)-Q^o(p+hu))K(u)du]^2\}$$
$$+\,\{\int_{-c}^{c}[Q^o(p+hu)-Q^o(p)]K(u)du\}^2$$
$$+\,2E\{\int_{-c}^{c}[Q^o(p+hu)-Q^o(p)]K(u)du\cdot\int_{-c}^{c}[\hat{Q}_n(p+hu)$$

$$-Q^{o}(p+hu)]K(u)du\}. \quad (4.1)$$

By the assumption that Q^{o} has bounded second derivative on some neighborhood of p,

$$\{\int_{-c}^{c}[Q^{o}(p+hu)-Q^{o}(p)]K(u)du\}^{2} = O(h^{2}). \quad (4.2)$$

For $\varepsilon > 0$ define the events $A_n = \{|U_n(p+hu)-(p+hu)| > \varepsilon\}$ where U_n is the PL quantile function based on the uniform distribution on (0,1) as in the proof of Theorem 3.1. By the same argument in that proof, choosing h small such that $hc < \min\{1, T_{H(Q^{o})}\}$, we have $P(A_n) \leq d_o \exp(-nd_1)$ for some positive constants d_o and d_1. Now write

$$E\{[\int_{-c}^{c}(\hat{Q}_n(p+hu)-Q^{o}(p+hu))K(u)du]^{2}\} = E_1 + E_2,$$

where

$$E_1 = E\{[\int_{-c}^{c}(\hat{Q}_n(p+hu)-Q^{o}(p+hu))K(u)du] \ . \ I_{A_n^c}\}.$$

and

$$E_2 = E\{[\int_{-c}^{c}(\hat{Q}_n(p+hu)-Q^{o}(p+hu))K(u)du] \ . \ I_{A_n}\}.$$

By the same argument as in the proof of Theorem 3.1, since $p+hu < \min\{1, T_{H(Q^{o})}\}$, $|E_2| = O(\exp(-nd_1'))$. Applying Taylor's formula to E_1 and using Sander's (1975) inequality (the symmetry property) gives

$$E_1 = E\{[\int_{-c}^{c} K(x)(U_n(p+hu)-(p+hu))Q^{o\prime}(p+hu)du]^{2}\}$$
$$+ O(E[\sup_{0\leq p\leq T^{*}} |\hat{U}_n(p)-p|^{3}]) + O(\exp(-nd_1')),$$

where $p < p+hc < T^{*} < \min\{1,T_{H(Q^{o})}\}$. From the proof of Theorem 2 of Lio, Padgett and Yu (1986) for large n, $O(E[\sup_{0\leq p\leq T^{*}} |\hat{U}_n(p)-p|^{3}]) = O(n^{-3/2})$. Also, by the same argument as in the proof of Theorem 3.1,

$$E\{[\int_{-c}^{c}K(x)(U_n(p+hx)-(p+hx))Q^{o\prime}(p+hx)dx]^{2}\}$$
$$= n^{-1}E\{[\int_{-c}^{c}K(x)[n^{\frac{1}{2}}(U_n(p+hx)-(p+hx))$$
$$- n^{-\frac{1}{2}}K^{*}(p+hx,n)]dx]^{2}\}(Q^{o\prime}(p))^{2}$$
$$+ 2n^{-1}E\{[\int_{-c}^{c}K(x)[n^{\frac{1}{2}}(U_n(p+hx)-(p+hx))$$
$$- n^{-\frac{1}{2}}K^{*}(p+hx,n)]dx[\int_{-c}^{c}K(x)n^{-\frac{1}{2}}K^{*}(p+hx,n)dx]]\}(Q^{o\prime}(p))^{2}$$

$$+ n^{-1}(Q^{o\prime}(p))^2 E\{[\int_{-c}^{c} K(x) n^{-\frac{1}{2}} K^*(p+hx,n)dx]^2\}$$

$$+O(n^{-3/2}) + o(hn^{-1})$$

$$= o(n^{-1}) + n^{-1}(Q^{o\prime}(p))^2 E\{[\int_{-c}^{c} K(x) n^{-\frac{1}{2}} K^*(p+hx,n)dx]^2\}$$

$$+O(n^{-3/2}) + o(hn^{-1}).$$

Now, by a result of Aly, Csörgö and Horváth (1985),

$$E\{[\int_{-c}^{c} K(x) n^{-\frac{1}{2}} K^*(p+hx,n)dx]^2\}$$

$$= E\{ \int_{-c}^{c} n^{-1}(1-p-hx)W(d(p+hx),n)K(x)dx$$

$$\times \int_{-c}^{c} (1-p-ht)W(d(p+ht),n)K(t)dt\}$$

$$= A_1 + A_2 + A_3,$$

where

$$A_1 = n^{-1}E\{ \int_{-c}^{c}\int_{-c}^{c}(1-p)^2 W(d(p+hx),n)W(d(p+ht),n)$$

$$\times K(x)K(t)dxdt\},$$

$$A_2 = -2n^{-1}h(1-p)\ E\{\int_{-c}^{c}\int_{-c}^{c} x\ K(x)K(t)\ W(d(p+hx),n)$$

$$\times W(d(p+ht),n)dxdt\},$$

$$A_3 = n^{-1}h^2 E\{\int_{-c}^{c}\int_{-c}^{c} xtK(x)K(t)W(d(p+hx),n)$$

$$\times W(d(p+ht),n)dxdt\},$$

and $W(s,t)$ denotes a two-parameter Weiner process with $E[W(s,t)]=0$ and $E[W(s,t)W(s',t')] = \min\{s,s'\} \min\{t,t'\}$ with $d(t) = \int_{-\infty}^{t}(1-x)^{-2}[1-H(Q^o(x))]^{-1}dx$.

Now,

$$A_1 = n^{-1}(1-p)^2\int_{-c}^{c}\int_{-c}^{t} K(t)K(u)\int_{0}^{p+hu}(1-x)^{-2}[1-H(Q^o(x))]^{-1}dx\ du\ dt$$

$$+ n^{-1}(1-p)^2\int_{-c}^{c}\int_{t}^{c} K(t)K(u)\int_{0}^{p+hu}(1-x)^{-2}[1-H(Q^o(x))]^{-1}dx\ du\ dt$$

$$= n^{-1}(1-p)^2\{\int_{-c}^{c}\int_{-c}^{t} K(t)K(u)\int_{0}^{p}(1-x)^{-2}[1-H(Q^o(x))]^{-1}dx\ du\ dt$$

$$+ \int_{-c}^{c}\int_{-c}^{t} K(t)K(u)\int_{p}^{p+hu}(1-x)^{-2}[1-H(Q^o(x))]^{-1}dx\ du\ dt$$

$$+ \int_{-c}^{c} K(t)[1-\tilde{K}(t)]\int_{0}^{p}(1-x)^{-2}[1-H(Q^o(x))]^{-1}dx\ dt$$

$$+ \int_{-c}^{c} K(t)[1-\tilde{K}(t)]\int_{p}^{p+ht}(1-x)^{-2}[1-H(Q\ (x))]^{-1}dx\ dt\}.$$

Combine the first and third terms in the last expression for A_1,

and in the second term let $g(u) = \int_p^{p+hu} (1-x)^{-2}[1-H(Q^o(x))]^{-1}dx$ and change the order of integration. Then combine the second and fourth terms to get

$$A_1 = n^{-1}(1-p)^2 \int_0^p (1-x)^{-2}[1-H(Q^o(x))]^{-1}dx$$
$$+ 2n^{-1}(1-p)^2 \int_{-c}^{c} K(t)[1-\tilde{K}(t)] \int_p^{p+ht} (1-x)^{-2}[1-H(Q^o(x))]^{-1}dx\ dt.$$

By the same arguments, A_2 and A_3 become

$$A_2 = -2\ n^{-1}h(1-p)\int_{-c}^{c} tK(t)\int_{-c}^{t} K(u)\int_p^{p+hu} (1-x)^{-2}[1-H(Q^o(x))]^{-1}dx\ du\ dt$$
$$-4\ n^{-1}h(1-p)\int_{-c}^{c} tK(t)[1-\tilde{K}(t)]\int_p^{p+ht} (1-x)^{-2}[1-H(Q^o(x))]^{-1}dx\ dt$$

and

$$A_3 = 2h^2\int_{-c}^{c} tK(t)\int_p^{p+ht} (1-x)^{-2}[1-H(Q^o(x))]^{-1}dx\left(\int_t^c sK(s)ds\right)dt.$$

Finally, combining these results for E_1 and the result for E_2, (4.1) yields the asymptotic expression of the theorem.///

Define $Q^o(p,h) = h^{-1}\int_0^1 Q^o(t)K((t-p)/h)dt$. Then an asymptotic expression for $E\{[Q_n(p)-Q^o(p,h)]^2\}$ can be obtained similar to that in Theorem 4.1.

<u>Theorem 4.2</u> With the same hypotheses as in Theorem 4.1, for $0 < h < \delta$ with small enough $\delta < 1$,

$$E\{[Q_n(p) - Q^o(p,h)]^2\}$$
$$= n^{-1}(1-p)^2(Q^{o\prime}(p))^2\int_0^p (1-x)^{-2}[1-H(Q^o(x))]^{-1}dx$$
$$+ 2n^{-1}(1-p)^2(Q^{o\prime}(p))^2\int_{-c}^{c} K(t)[1-\tilde{K}(t)]$$
$$\times \int_p^{p+ht} (1-x)^{-2}[1-H(Q^o(x))]^{-1}dx\ dt$$
$$+ O(n^{-3/2}) + O(h^2n^{-1}) + o(hn^{-1}) + o(n^{-1}).$$

Note that for h sufficiently small, we have $E\{[Q_n(p)-Q^o(p)]^2\} = E\{[Q_n(p)-Q^o(p,h)]^2\} + O(h^2)$. Hence, the two expectations are close for large n and small h. A

comparison of the mean squared error of the PL quantile function with the result of Theorem 4.2 can be stated in the following corollary. The condition on the kernel function in this corollary is the same condition as in Falk (1984).

Corollary 4.3 If $\int_{-c}^{c} tK(t)\tilde{K}(t)\,dt > 0$, then under the conditions of Theorems 3.1 and 4.2, there exists a $\delta > 0$ such that for any fixed bandwidth $0 < h < \delta$ there is an N_o so that when $n > N_o$,
$E\{[Q_n(p) - Q^o(p,h)]^2\} - E\{[\hat{Q}_n(p)-Q^o(p)]^2\} < 0$.

Proof: Write

$$\frac{n}{h}\left[E\{[Q_n(p)-Q^o(p,h)]^2\} - E\{[\hat{Q}_n(p)-Q^o(p)]^2\}\right]$$
$$= 2h^{-1}(1-p)^2(Q^{o\prime}(p))^2 \int_{-c}^{c} K(t)[1-\tilde{K}(t)]$$
$$\times\int_p^{p+ht}(1-x)^{-2}[1-H(Q^o(x))]^{-1}dx\,dt$$
$$+ O(n^{-\frac{1}{2}}h^{-1}) + o(1) + h^{-1}o(1) + O(h)$$

which for large n and small h is approximately

$$-2(Q^{o\prime}(p))^2 \int_{-c}^{c} tK(t)\tilde{K}(t)dt[1-H(Q^o(p))]^{-1} < 0.///$$

Remarks: An attempt to extend Falk's (1985) methods for kernel type quantile estimators to the case of random right-censorship in a straightforward manner presents difficult mathematical problems. In order to obtain a direct comparison of the mean squared error of $\hat{Q}_n(p)$ with that of $Q_n(p)$, a rate of convergence faster than the $o(n^{-1})$ term in the expression in Theorem 4.1 is needed. However, such a rate is not available. A relationship between the rates at which $h \to 0$ and $n \to \infty$ seems to be required to determine the relative behavior of these two estimators with respect to their mean squared errors.

ACKNOWLEDGEMENTS

This research was supported by the U.S. Air Force Office of Scientific Research grant number AFOSR-84-0156 and the U.S. Army Research Office grant number MIPR-ARO-139-85.

REFERENCES

Aly, E.-E.A.A., Csörgő, M. and Horváth, L. (1985). Strong Approximations of the Quantile Process of the Product-Limit Estimator. Journal of Multivariate Analysis, 16, 185-210.

Cheng, K.F. (1984). On Almost Sure Representations for Quantiles of the Product Limit Estimator with Applications. Sankhya, Ser.A 46, 426-443.

Csörgő, M. (1983). Quantile Processes with Statistical Applications. CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, PA.

Efron, B. (1967). The Two-Sample Problem with Censored Data. Proceedings of the Fifth Berkeley Symposium, 4, 831-853.

Falk, M. (1984). Relative Deficiency of Kernel Type Estimators of Quantiles. Annals of Statistics, 12, 261-268.

Falk, M. (1985). Asymptotic Normality of the Kernel Quantile Estimator. Annals of Statistics, 13, 428-433.

Földes, A., and Rejtö, L. (1981). A LIL Type Result for the Product Limit Estimator. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwande Gebiete, 56, 76-86.

Kaplan, E. L., and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association, 53, 457-481.

Lio, Y. L. and Padgett, W. J. (1986). Some Convergence Results for Kernel-Type Quantile Estimators Under Censoring. Statistics and Probability Letters, (in press).

Lio, Y. L., Padgett, W. J., and Yu, K. F. (1986). On the Asymptotic Properties of a Kernel-Type Quantile Estimator from Censored Samples. Journal of Statistical Planning and Inference, 14, 169-177.

Padgett, W. J. (1986). A Kernel-Type Estimator of a Quantile Function from Right-Censored Data. Journal of the American Statistical Association, 81, 215-222.

Padgett, W. J. and Thombs, L. A. (1986). Smooth Nonparametric Quantile Estimation Under Censoring: Simulations and Bootstrap Methods. Communications in Statistics - Simulation and Computation, (in press).

Parzen, E. (1979). Nonparametric Statistical Data Modeling. Journal of the American Statistical Association, 74, 105-121.

Sander, J. (1975). The Weak Convergence of Quantiles of the Product Limit Estimator. Technical Report 5, Stanford University, Department of Statistics.

Yang, S. S. (1985). A Smooth Nonparametric Estimator of a Quantile Function. Journal of the American Statistical Association, 80, 1004-1011.

END

12-86

DTIC